12-2020

Medical Knowledge-enriched Textual Entailment Framework

Shweta Yadav

Vishal Pallagani

Amit P. Sheth
University of South Carolina - Columbia, amit@sc.edu

Publication Info

Preprint version *Proceedings of the 28th International Conference on Computational Linguistics*, 2020.

Medical Knowledge-enriched Textual Entailment Framework

Shweta Yadav[*], **Vishal Pallagani**[‡], **Amit Sheth**[‡]
[*]LHNCBC, U.S. National Library of Medicine, MD, USA
[‡] University of South Carolina, SC, USA
[*]shweta.shweta@nih.gov, [‡]{VISHAL,AMIT}@sc.edu

Abstract

One of the cardinal tasks in achieving robust medical question answering systems is textual entailment. The existing approaches make use of an ensemble of pre-trained language models or data augmentation, often to clock higher numbers on the validation metrics. However, two major shortcomings impede higher success in identifying entailment: **(1)** understanding the focus/intent of the question and **(2)** ability to utilize the real-world background knowledge to capture the context beyond the sentence. In this paper, we present a novel Medical Knowledge-Enriched Textual Entailment framework that allows the model to acquire a semantic and global representation of the input medical text with the help of a relevant domain-specific knowledge graph. We evaluate our framework on the benchmark MEDIQA-RQE dataset and manifest that the use of knowledge-enriched dual-encoding mechanism help in achieving an absolute improvement of 8.27% over SOTA language models. We have made the source code available here.[1]

1 Introduction

The entailment task is similar with natural language inference (NLI), involves identifying the semantic similarity between two natural language texts, premise (P) and hypothesis (H). The NLI task's effectiveness is crucial for developing a robust natural language understanding system that functions at a human level (Ben Abacha et al., 2019; Abacha and Demner-Fushman, 2016; Romanov and Shivade, 2018). Recent literature suggests the use of contemporary language models (LMs) (Devlin et al., 2018; Beltagy et al., 2019), often ensembled, to achieve better performance (Zhu et al., 2019; Bhaskar et al., 2019; Xu et al., 2019) on the NLI task. However, our qualitative interpretation of the dataset and results suggests that LMs fails when it comes to textual entailment (TE), despite being the on-demand language model. The limitations belong to two major categories:

- **Multiple word form:** Medical text offers high degree of variability in the form of synonym and abbreviated words. The same can be witnessed in Table-1, where BERT (Devlin et al., 2018) is unable to predict the entailment between P and H as they have different terms - *'Kartagener's Syndrome'* and *'Primary Ciliary Dyskinesia'*, while both are synonym.
- **Focus/Intent understanding:** Given P and H, LMs often fails to capture the focus/intent of both the sentences. Table-1 shows an example, where the focus of P and H are misunderstood. It can be seen that P emphasizes the possibility of *'atypical pneumonia'* occurring within a month after treatment, whereas H talks about the possible treatments for the disease.

These findings indicate that existing LMs lack semantic interpretation of the input, which is crucial in the inferencing tasks. In this paper, we deal with the question:

Does the medical textual entailment task benefit from the external domain knowledge to distinguish semantically identical medical sentences to recognize entailment?

To address this question and above-mentioned limitations, this paper presents a novel framework for

[1]https://github.com/VishalPallagani/Medical-Knowledge-enriched-Textual-Entailment

entails	**Premise:** I am suffering from Kartagener's syndrome and wanted information from you or from Dr. [NAME] for this syndrome. (About fertility) and if possible other symptoms. **Hypothesis:** *What is primary ciliary dyskinesia?*
not entails	**Premise:** What is the possibility of atypical pneumonia occurring again less than a month after treatment? **Hypothesis:** *What are the possible treatments for atypical pneumonia?*

Table 1: Examples from the MedQA-RQE dataset, where the text highlighted in blue are semantically similar words and red represents the lexically similar but semantically dissimilar words.

recognizing textual entailment, `Sem-KGN`: **Sem**antic **K**nowledge-enriched **G**raph **N**etwork that explores the domain-specific knowledge to enhance the semantic interpretability in the LMs. The proposed method devise a dual-encoding mechanism to enrich the classical document encoding (obtained from BERT) with the knowledge-enriched graph encoder. Specifically, our method builds a heterogeneous dictionary graph of the given P and H to encode the global context, while BERT's proficiency lies in capturing the local contextual information. Rather than constructing graphs solely based on the triples (subject, object, predicate), more *semantic units* are introduced into the graph as additional nodes to enrich the relations between the entities. These additional nodes add medical-entities centered factual information which are generated by expanding the medical knowledge graphs (KGs) such as UMLS (Bodenreider, 2004), SNOMED-CT (Donnelly, 2006) and ICD10 (Quan et al., 2005). The medical entities present in P and H are used to query the related information, *'diseases/syndromes'*, *dosage*, *'side-effects'*, and *'drug-interaction'* from the mentioned KGs. Finally, Graph Convolutional Network (GCN) (Kipf and Welling, 2017) is employed to generate the graph encoding, augmented with the regular document encoder, and later fused through a multi-headed attention layer to support entailment.

Contributions: **(i)** Proposed medical-TE framework, by utilizing domain-specific medical KGs to encode the global and semantic information of the premise and hypothesis, **(ii)** Exploited the capabilities of semantic units in a graph network to encode medical-entities centered factual information for recognizing textual entailment, and **(iii)** Evaluated the effectiveness of the proposed method over state-of-the-art language models and knowledge-infused baseline methods on the benchmark MEDIQA-RQE dataset.

Related Work : The development of annotated TE and NLI medical datasets (Abacha et al., 2015; Ben Abacha et al., 2019; Abacha and Demner-Fushman, 2016; Romanov and Shivade, 2018) and a variety of pre-trained language models has led to a rise of extensive ongoing research in this field. Majority of the systems developed for the TE task adopts the multi-task learning (MTL) framework (Zhu et al., 2019; Bhaskar et al., 2019; Kumar et al., 2019; Zhou et al., 2019; Xu et al., 2019), ensemble method (Sharma and Roychowdhury, 2019), and transfer learning (Bhaskar et al., 2019) for achieving better accuracy. Xu et al. (2019) employed the MTL approach (Liu et al., 2019; Yadav et al., 2018; Yadav et al., 2019; Yadav et al., 2020) in TE task to learn from the auxiliary tasks of question answering (QA) and NLI. The best performing system at MedQA 2019-RQE shared task (Zhu et al., 2019) utilized the MTL approach to learn from intermediate NLI task. Further they used knowledge distillation approach to condense the information obtained from various models and transfer it into an single model. Few of the works (Wang et al., 2019; Khot et al., 2018) have explored the usage of the background knowledge or medical KGs (Kumar et al., 2019; Bhaskar et al., 2019) in extracting information for the entailment task. However, the consideration of building a vocabulary graph from the textual and later enriching them with information from the medical KGs is still an unexplored territory.

2 Proposed Approach

The overall architecture of the proposed `Sem-KGN` is illustrated in Fig.-1. The rest of the section elaborates `Sem-KGN` in detail.

2.1 Document Encoder

Given a n_P words premise $P = \{w_1^P, w_2^P, \ldots, w_{n_P}^P\}$ and n_H words hypothesis $H = \{w_1^H, w_2^H, \ldots, w_{n_H}^H\}$. The document encoder is responsible to capture basic lexical and syntactic in-

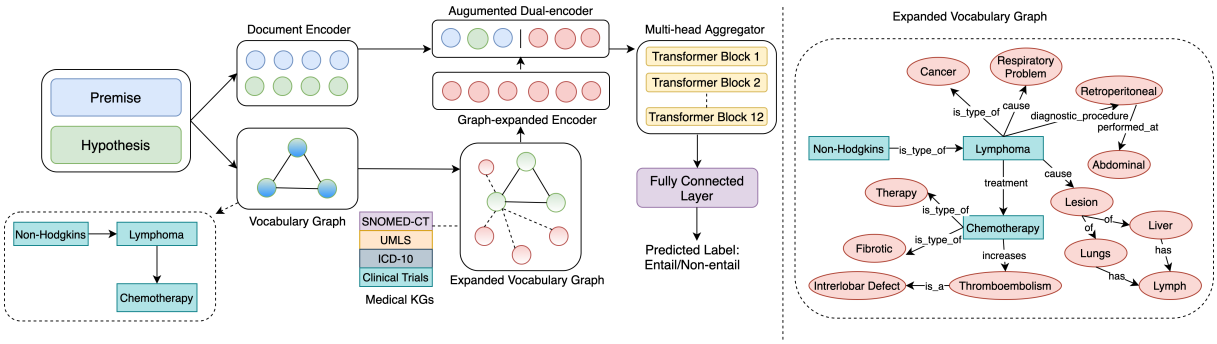

Figure 1: Architecture of proposed methodology

formation from the premise and hypothesis in the form of local context information. We employed the BERT model to serve as the document encoder in our proposed `Sem-KGN` framework. Formally, the local context features are computed as $d_1, d_2, \ldots, d_n = \texttt{Document-Encoder}(w_1, w_2, \ldots, w_n)$.

2.2 Knowledge-enriched Graph Encoder

Vocabulary Graph Construction: We first construct a dictionary based on all the unique words in the training dataset. Thereafter, we build a graph $G = (V, E)$ based on the word co-occurrence information in the dictionary. Instead of building graph based on a given P and H, we were motivated by the work of (Lu et al., 2020) to build graph by considering all the lexicon in the dataset, which aims at encoding the global information of the particular domain (in our case medical domain). The nodes of the graph G are words in the dictionary, the edge between two nodes w_i and w_j is determine by the normalized point-wise mutual information (NPMI) (Bouma, 2009).

$$\texttt{NPMI}(w_i, w_j) = -ln\frac{p(w_i, w_j)}{p(w_i)p(w_j)}\frac{1}{ln\ p(w_i, w_j)} \tag{1}$$

where $p(w_i, w_j) = \frac{\#s(w_i, w_j)}{\#W}$, $p(i) = \frac{\#s(w_i)}{\#W}$, $\#s(.)$ is the number of sliding windows containing a word or a pair of words, and $\#W$ is the total number of sliding windows. We make an edge between the nodes if the value of $\texttt{NPMI}(w_i, w_j)$ exceed a particular threshold value.

Graph-expansion with Medical Knowledge: We expand the existing graph G with the additional nodes and corresponding edges to form a graph $G^* = (V \cup \overline{V}, E \cup \overline{E})$. Towards this, first we extracted the medical entities by exploiting the entity recognition model[2] trained on MedMentions dataset (Mohan and Li, 2019), from a given pair P and H. Once the medical entities are identified, KGs such as SNOMED-CT, ICD-10, UMLS and Clinical Trials are exploited to extract the information of two-hop connected *'diseases/syndromes'*, *'dosage'*, *'side effects'*, and *'drug-interaction'* type medical-concepts. The final expanded entities are the additional nodes $\overline{V}$ which act as the *semantic unit* to capture the domain-specific relationship (e.g., *treats*, *caused by*) and hierarchical relations (e.g., *is a*) between medical-concepts.

Graph-expanded Knowledge Encoding: The **G**raph-**e**xpanded **K**nowledge **Encoder** (`GxK Encoder`) computes the representation of each node from graph G^*. We are interested to compute the representation for each extracted entities $(e_1, e_2 \ldots e_m)$ for given pair of P and H. Formally, we get $h_1, h_2, \ldots, h_m = \texttt{GxK-Encoder}(G^*, e_1, e_2, \ldots, e_m)$

We model it using the 2-layer GCN architecture. For the entities, we first make a input matrix $M \in \mathbb{R}^{m \times |V|}$, where row of the matrix M is the one-hot vector of length of $|V|$ (size of dictionary). Given the adjacency matrix D of expanded graph G^* a and input matrix M, a single layer of graph convolution is computed as follows:

$$H^{(1)} = relu(MDW^{(1)}),\ \text{where } D \in \mathbb{R}^{|V| \times |V|} \text{ and } W^{(1)} \in \mathbb{R}^{|V| \times d} \tag{2}$$

[2] `https://go.aws/37SD7ae`

where $H^{(1)} \in \mathbb{R}^{m \times d}$ is matrix which rows are the node features in. For a given input M, we are interested to captures the part of the graph from D by multiplying them together as MD. The feature at a given node w_i computed by the interaction between all the neighbouring nodes as $h_{w_i}^{(l+1)} = relu\left(\sum_j \frac{1}{c_{ij}} h_{w_j}^{(l)} W^{(l)}\right)$ where c_{ij} is the normalization constant (Kipf and Welling, 2017). The second layer convolution is obtained as $H^{(2)} = relu(H^{(1)}W^{(2)})$, where $W^{(2)} \in \mathbb{R}^{d \times d}$

2.3 Multi-headed Aggregator

We fuse the information from `Document-Encoder` and `GxK-Encoder` using the multi-headed self-attention (Vaswani et al., 2017). Our aims to utilize the best of both the worlds. We form a single feature sequence by augmenting both the encoder representations obtained from both the encoders, with additional `[CLS]` token representation. By applying the self-attention on the augmented encoding sequence, we facilitate the network to attend the useful information across the individual encoding. We model our **M**ulti-**H**eaded **Aggregator** (`MH-Aggregator`) using the 12 layers of Transformer-block with 16 heads. We use the last layer output of the aggregator and consider the `[CLS]` token representation as the final representation $F \in \mathbb{R}^d$ of given P and H pair. We employ a feed-forward layer to classify a pair of premise P and hypothesis H into the corresponding *'entail'* or *'non-entail'* classes.

$$
\begin{aligned}
f_1, f_2, \ldots, f_{n+m} &= \texttt{MH-Aggregator}(d_1, d_2, \ldots, d_n, h_1, h_2, \ldots, h_m) \\
prob(class = entail|P, H, \theta) &= exp(W_{entail}^T F + b) / \sum_j exp(W_j^T F + b)
\end{aligned} \tag{3}
$$

Models		Accuracy	Precision	Recall	F1-Score
Baseline 1	BERT	47.90	46.16	48.26	47.18
	BioBERT	50.14	46.28	48.97	47.58
	ClinicalBERT	49.60	48.79	49.56	49.17
Baseline 2	BERT + KI	49.56	46.06	48.69	47.33
	BioBERT + KI	51.15	48.56	49.56	49.05
	ClinicalBERT + KI	50.14	50.00	50.00	50.00
(Zhu et al., 2019)	BERT + linear projection	51.30	51.53	51.30	49.45
Proposed Model	`Sem-KGN`	**56.17**	**63.18**	**56.18**	**59.47**

Table 2: Experimental results of our proposed model (`Sem-KGN`) and the baseline methods on the official test set.

Models Components	Accuracy	Precision	Recall	F1-Score
`Sem-KGN`	56.17	63.18	56.18	59.47
(-) Knowledge-enriched Graph Encoder	47.90 (8.27 ↓)	46.16 (17.02 ↓)	48.26 (7.90 ↓)	47.18 (12.29 ↓)
(-) Medical Knowledge-graph	50.65 (5.52 ↓)	62.27 (0.91 ↓)	50.66 (5.52 ↓)	55.86 (3.61 ↓)

Table 3: Ablation study showing the role of each component in the model. The values within the bracket show the absolute decrements by removing the component.

3 Experimental Results and Analysis

Dataset and Metrics: We used widely adopted benchmark entailment dataset, MEDIQA-RQE created by (Abacha and Demner-Fushman, 2016), released in the BioNLP 2019 shared task. The dataset is derived from consumer health questions (CHQs) and frequently asked questions (FAQs) from the U.S. National Library of Medicine and National Institute of Health respectively. The training and validation set consists of total 8890 CHQ and FAQ pairs with the entail label of 4784 instances. The test set consist of 230 pairs with 115 entail labels. We used official evaluation metrics (Accuracy) to evaluate our model. Additionally, we also provided the Precision, Recall, and F1-Scores for the evaluation.

Implementation Details: We have chosen models' hyper-parameters empirically on the validation set. The base-uncased version of BERT[3] of hidden size 768 with a max sequence length of 200 (160 for P and 40 for H) is used in all experiments reported in the paper. The size of dictionary to create the dictionary matrix was 30000. The dimension of The threshold of NPMI is set to 0.3 to obtain meaningful relation between words. The last layer's hidden size of the graph-expanded knowledge encoder is set to 16. We use the Adam optimiser (Kingma and Ba, 2014) for parameters update after every epoch of training. We set the standard batch size of 16, and trained for 5 epochs with a dropout rate of 0.2 and $2e-5$ learning rate in all the experimental results reported in this work.

Baseline Models: To show the effectiveness of `Sem-KGN`, we adopted following competitive baseline models:

1. Language Models (LMs): We utilized the SOTA LMs (*BERT*) as well as the LMs adapted for the

[3] `https://bit.ly/2ZajZR1`

Sample Examples and Error Type	Premise	Hypothesis	True Label	BioBERT+ KI	Sem-KGN
Example 1	*Can you mail me patient information about Glaucoma, I was recently diagnosed and want to learn all I can about the disease.*	*How is glaucoma diagnosed ?*	Non-entailed	Non-entailed	Non-entailed
Example 2	*I was writing to inquire about more information regarding the diagnosis of OI. We have family members who are in the process of waiting for genetic testing to come back but are under allegations of child abuse. Is there any information that may be helpful to us?*	*How to diagnose Osteogenesis Imperfecta ?*	Entailed	Non-entailed	Entailed
Error Type 1: Complex and Long Question	*Hello, my dad, 68 years old, has gastritis, it did ache occasionally over the last several years. The other day, he went to hospital to have medical check-up with endoscopic ultrasonography, and found GIST with about 1cm in size. Dr. told him that he may consider surgery or not, it is up to him. What are we supposed to do?*	*How is an endoscopic ultrasound performed ?*	Non-entailed	Entailed	Non-entailed
Error Type 2: Ambiguous Question	*Can you please send me as much information as possible on "hypothyroidism". I was recently diagnosed with the disease and I am struggling to figure out what it is and how I got it.*	*How is Hypothyroidism diagnosed ?*	Non-entailed	Entailed	Non-entailed

Table 4: Qualitative and error analysis of our proposed model (Sem−KGN) with the best baseline model.

clinical (*ClinicalBERT*) and medical domain (*BioBERT*), fine-tuned for MEDIQA-RQE task.

2. Knowledge-Infused Language Models (+KI): These baselines model works on the principle of *shallow knowledge infused learning*, where we provided the knowledge about the medical entities at the token level to the LMs. The intuition behind using this baseline was to understand at what layer if knowledge is integrated into the LMs, it is going to be beneficial.

Results: Table-2 provides an overview of the results, which demonstrates that Sem−KGN, equipped with KGs enriched graph encoding performs the best over all the baselines model. A considerable increase in the accuracy of 8.27% can be observed over vanilla BERT model in comparison to proposed Sem−KGN. The similar set of improvement (over 6%) can be observed with BioBERT and ClinicalBERT. We also observed the power of basic shallow (+KI) over the vanilla LMs showing the absolute improvement of 1.5%. Further, in comparison to baseline 2, Sem−KGN achieved the average increment of 6% over all the knowledge-infused LMs. Finally, from our ablation study (*c.f.* Table-3), it can be noticed that enriching the graph encoder with the domain knowledge assists in the entailment task. The results conclude two important claims: **(1)** the modularity of the knowledge infusion process that can be combined with any LMs is witnessed, and **(2)** Sem−KGN proves its effectiveness in having a local as well global understanding of the premise and hypothesis. We also compare the results against the best system (Zhu et al., 2019) at MEDIQA-RQE task that was based on the ensemble of LMs. However, to have a fair comparison and understand the role of our Knowledge-enriched graph encoder, we only utilize their model that have introduced "linear projection" over BERT. The proposed Sem−KGN has outperform the "linear projection" mechanism described in Zhu et al. (2019).

Analysis: Table-4 depicts the qualitative analysis of Sem−KGN, w.r.t baseline models on 50 randomly sampled DEV set. The first entry shows the effectiveness of Sem-KGN in being able to understand the focus of the premise and the hypothesis which is achieved by the dual encoding mechanism. The second entry in the table affirms the importance of domain-specific KGs in assimilating medical information.

Error Analysis: Table-4 entry 3 and 4 shows the leading cause of the error in the proposed model. We found that major misclassification occurred when the premise was complex and have a multiple questions. We also observed in some cases when there is high ambiguity between premise and hypothesis, model fail to recognize the correct label. For e.g., in the Table-4 entry 4, with the presence of term *'hyperthyroidism'* and *'diagnosed'* both in premise and hypothesis it is very difficult to recognize the true label: not-entailed.

4 Conclusion

In this paper, we proposed a framework Sem−KGN to recognize medical textual entailment. Our framework utilized the local context from BERT based document encoder and global context by expanding the vocabulary graph with the medical entities obtained from the medical knowledge-bases. We present an efficient aggregator scheme to fuse the multiple encoding. The proposed Sem−KGN framework outperformed the competent pre-trained language model and knowledge-graph enabled language model architectures with fair margin. In future, we plan to explore multi-domain knowledge-graph and efficient graph embedding based techniques for medical textual entailment.

References

Asma Ben Abacha and Dina Demner-Fushman. 2016. Recognizing question entailment for medical question answering. In *AMIA Annual Symposium Proceedings*, volume 2016, page 310. American Medical Informatics Association.

Asma Ben Abacha, Duy Dinh, and Yassine Mrabet. 2015. Semantic analysis and automatic corpus construction for entailment recognition in medical texts. In *Conference on Artificial Intelligence in Medicine in Europe*, pages 238–242. Springer.

Iz Beltagy, Kyle Lo, and Arman Cohan. 2019. Scibert: A pretrained language model for scientific text. *arXiv preprint arXiv:1903.10676*.

Asma Ben Abacha, Chaitanya Shivade, and Dina Demner-Fushman. 2019. Overview of the mediqa 2019 shared task on textual inference, question entailment and question answering. In *ACL-BioNLP 2019*.

Sai Abishek Bhaskar, Rashi Rungta, James Route, Eric Nyberg, and Teruko Mitamura. 2019. Sieg at mediqa 2019: Multi-task neural ensemble for biomedical inference and entailment. In *Proceedings of the 18th BioNLP Workshop and Shared Task*, pages 462–470.

Olivier Bodenreider. 2004. The unified medical language system (umls): integrating biomedical terminology. *Nucleic acids research*, 32(suppl_1):D267–D270.

Gerlof Bouma. 2009. Normalized (pointwise) mutual information in collocation extraction. *Proceedings of GSCL*, pages 31–40.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Kevin Donnelly. 2006. Snomed-ct: The advanced terminology and coding system for ehealth. *Studies in health technology and informatics*, 121:279.

Tushar Khot, Ashish Sabharwal, and Peter Clark. 2018. Scitail: A textual entailment dataset from science question answering. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*.

Vinayshekhar Bannihatti Kumar, Ashwin Srinivasan, Aditi Chaudhary, James Route, Teruko Mitamura, and Eric Nyberg. 2019. Dr. quad at mediqa 2019: Towards textual inference and question entailment using contextualized representations. *arXiv preprint arXiv:1907.10136*.

Xiaodong Liu, Pengcheng He, Weizhu Chen, and Jianfeng Gao. 2019. Improving multi-task deep neural networks via knowledge distillation for natural language understanding. *arXiv preprint arXiv:1904.09482*.

Zhibin Lu, Pan Du, and Jian-Yun Nie. 2020. Vgcn-bert: Augmenting bert with graph embedding for text classification. In *European Conference on Information Retrieval*, pages 369–382. Springer.

Sunil Mohan and Donghui Li. 2019. Medmentions: a large biomedical corpus annotated with umls concepts. *arXiv preprint arXiv:1902.09476*.

Hude Quan, Vijaya Sundararajan, Patricia Halfon, Andrew Fong, Bernard Burnand, Jean-Christophe Luthi, L Duncan Saunders, Cynthia A Beck, Thomas E Feasby, and William A Ghali. 2005. Coding algorithms for defining comorbidities in icd-9-cm and icd-10 administrative data. *Medical care*, pages 1130–1139.

Alexey Romanov and Chaitanya Shivade. 2018. Lessons from natural language inference in the clinical domain. *arXiv preprint arXiv:1808.06752*.

Prakhar Sharma and Sumegh Roychowdhury. 2019. Iit-kgp at mediqa 2019: Recognizing question entailment using sci-bert stacked with a gradient boosting classifier. In *Proceedings of the 18th BioNLP Workshop and Shared Task*, pages 471–477.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.

Xiaoyan Wang, Pavan Kapanipathi, Ryan Musa, Mo Yu, Kartik Talamadupula, Ibrahim Abdelaziz, Maria Chang, Achille Fokoue, Bassem Makni, Nicholas Mattei, et al. 2019. Improving natural language inference using external knowledge in the science questions domain. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 7208–7215.

Yichong Xu, Xiaodong Liu, Chunyuan Li, Hoifung Poon, and Jianfeng Gao. 2019. Doubletransfer at mediqa 2019: Multi-source transfer learning for natural language understanding in the medical domain. *arXiv preprint arXiv:1906.04382*.

Shweta Yadav, Asif Ekbal, Sriparna Saha, Pushpak Bhattacharyya, and Amit Sheth. 2018. Multi-task learning framework for mining crowd intelligence towards clinical treatment.

Shweta Yadav, Asif Ekbal, Sriparna Saha, and Pushpak Bhattacharyya. 2019. A unified multi-task adversarial learning framework for pharmacovigilance mining. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5234–5245.

Shweta Yadav, Srivastsa Ramesh, Sriparna Saha, and Asif Ekbal. 2020. Relation extraction from biomedical and clinical text: Unified multitask learning framework. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.

Huiwei Zhou, Xuefei Li, Weihong Yao, Chengkun Lang, and Shixian Ning. 2019. Dut-nlp at mediqa 2019: an adversarial multi-task network to jointly model recognizing question entailment and question answering. In *Proceedings of the 18th BioNLP Workshop and Shared Task*, pages 437–445.

Wei Zhu, Xiaofeng Zhou, Keqiang Wang, Xun Luo, Xiepeng Li, Yuan Ni, and Guotong Xie. 2019. Panlp at mediqa 2019: Pre-trained language models, transfer learning and knowledge distillation. In *Proceedings of the 18th BioNLP Workshop and Shared Task*, pages 380–388.